DOCUMENT RESUME

ED 226 705 IR 010 582

AUTHOR Brown, Roger L.
TITLE The Use of Computer-Assisted Identification of ARIMA Time-Series.
PUB DATE Aug 82
NOTE 24p.; Paper presented at the Annual Meeting of the American Psychological Association (Washington, DC, August 1982).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Assisted Instruction; Computer Programs; *Data Analysis; Higher Education; *Instructional Materials; Mathematics Instruction; Models; *Research Methodology; Social Psychology; *Statistical Analysis; Teaching Methods
IDENTIFIERS Autocorrelation; *Box Jenkins Forecasting Model; Time Series Analysis; Tutorial Mode

ABSTRACT

This study was conducted to determine the effects of using various levels of tutorial statistical software for the tentative identification of nonseasonal ARIMA models, a statistical technique proposed by Box and Jenkins for the interpretation of time-series data. The Box-Jenkins approach is an iterative process encompassing several stages of development, with the initial step requiring a tentative identification which relies on two statistics, the autocorrelation function (AUCF) and the partial autocorrelation function (PACF). Both functions are calculated by the statistical software on the time-series data and supplied to the user as a correlogram. A total of 56 time-series data files were evaluated by 30 novice undergraduate students in a repeated measures procedure. Students were randomly assigned to statistical software packages which produced either AUCF and PACF correlograms (a non-tutorial package), the correlograms and access to theoretical ARIMA model correlogram examples, or correlograms and access to a decision-support tutorial. Results indicated that the use of tutorial software significantly improved the identification of ARIMA models over non-tutorial software, and the type of tutorial significantly interacted with the degree of complexity of the time-series data. The report includes 4 references, and 10 tables and figures presenting study data. (LMM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED226705

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
  Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

THE USE OF COMPUTER-ASSISTED IDENTIFICATION OF ARIMA TIME-SERIES

by

Roger L. Brown

Center for Research on Pre-Traumatic Behavior

University of Wisconsin

Paper presented at the meeting of the American Psychological Association, Washington, D. C., August, 1982.

IR010582

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Roger L. Brown

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

THE USE OF COMPUTER-ASSISTED IDENTIFICATION OF ARIMA TIME-SERIES

ABSTRACT

BY

R. L. BROWN, PH.D.

UNIVERSITY OF WISCONSIN

THIRTY NOVICE UNDERGRADUATE STUDENTS WERE USED TO EVALUATE THE EFFECTIVENESS OF TWO LEVELS OF STATISTICAL TUTORIAL SOFTWARE USED IN THE IDENTIFICATION OF NONSEASONAL ARIMA MODELS. A TOTAL OF 56 TIME-SERIES DATA FILES WERE EVALUATED BY THE STUDENTS IN A REPEATED MEASURES FASHION. RESULTS INDICATED THAT THE USE OF TUTORIAL SOFTWARE SIGNIFICANTLY IMPROVED THE IDENTIFICATION OF ARIMA MODELS OVER NON-TUTORIAL SOFTWARE. IT WAS ALSO INDICATED THAT THE TYPE OF TUTORIAL SIGNFICANTLY INTERACTED WITH THE DEGREE OF COMPLEXITY OF THE TIME-SERIES DATA.

INTRODUCTION

THE NEED FOR ACCURATE STATISTICAL TECHNIQUES IN THE INTERPRETATION OF TIME-SERIES DATA IS CRUCIAL FOR MANY SOCIAL-PSYCHOLOGICAL STUDIES. WHILE MANY TECHNIQUES HAVE BEEN PROPOSED AND DEBATED, THE TECHNIQUE PROPOSED BY BOX AND JENKINS (1970) KNOWN AS ARIMA MODELING SEEMS TO SHOW PROMISE. WHILE MUCH HAS BEEN WRITTEN ABOUT THE TECHNIQUE, THERE SEEMS TO BE SOME RELUCTANCE ON THE PART OF THE SOCIAL-PSYCHOLOGICAL COMMUNITY TO USE IT. THE REASONS FOR THIS RELUCTANCE HAVE BEEN DISCUSSED ELSEWHERE (MC CLEARY AND HAY, 1980; BROWN, 1980). THIS PAPER EXPLORES ONE ASPECT OF THIS RELUCTANCE, THE DIFFICULTY OF IDENTIFYING A TENTATIVE ARIMA MODEL.

THE BOX-JENKINS APPROACH IS AN ITERATIVE PROCESS ENCOMPASSING SEVERAL STAGES OF DEVELOPMENT (SEE FIGURE 1), WITH THE INITIAL STEP REQUIRING A TENTATIVE IDENTIFICATION. THIS MODEL IDENTIFICATION STEP BASICALLY RELIES ON TWO STATISTICS: THE AUTOCORRELATION FUNCTION (AUCF) AND A MORE COMPLEX FUNCTION KNOWN AS THE PARTIAL AUTOCORRELATION FUNCTION (PACF). AUTOCORRELATION REFERS TO THE CORRELATIONS AMONG SUCCESSIVE DATA POINTS SEPARATED BY EACH DIFFERENT OBSERVATION (OR LAG) IN THE SERIES. PARTIAL AUTOCORRELATION REFERS TO THE AUTOCORRELATION OF A SERIES WITH THE INFLUENCE OF PREVIOUS TIME POINTS "PARTIALED OUT." THESE TWO FUNCTIONS ARE CALCULATED BY THE STATISTICAL SOFTWARE ON THE TIME-SERIES DATA AND SUPPLIED TO THE USER AS A CORRELOGRAM,

A PLOT OF THE AUTOCORRELATION AND PARTIAL AUTOCORRELATION AS A FUNCTION OF K LAGS (SEE FIGURE 2).

IT IS FROM AN OCCULAR ANALYSIS OF THE CORRELOGRAMS THAT A TENTATIVE ARIMA MODEL MAY BE IDENTIFIED. THIS OCCULAR ANALYTIC TECHNIQUE BASICALLY COMPARES THE ESTIMATED AUCF AND PACF WITH THEORETICAL OR EXPECTED FUNCTIONS, ACCORDING TO A VARIETY OF BASIC ARIMA MODELS. AMBIGUITY IN THESE ESTIMATED FUNCTIONS IS THEN LESSENED BY EITHER PLACING AN APPROXIMATE 95 PERCENT CONFIDENCE BAND AROUND EACH FUNCTION (±2 STANDARD ERRORS), OR BY CALCULATING A T-LIKE STATISTIC (BOWERMAN AND O'CONNEL, 1979). THE T-LIKE STATISTIC, FOR BOTH FUNCTIONS, IS SIMPLY THE VALUE OF EACH AUCF AND PACF OVER THEIR RESPECTIVE STANDARD ERRORS AT EACH LAG PERIOD.

THE PURPOSE OF THIS STUDY WAS TO INVESTIGATE THE EFFECTS OF USING VARIOUS LEVELS OF TUTORIAL STATISTICAL SOFTWARE FOR THE TENTATIVE IDENTIFICATION OF NONSEASONAL ARIMA MODELS.

METHODS

THIRTY NOVICE UNDERGRADUATE STUDENTS UNFAMILIAR WITH ARIMA MODELING WERE RECRUITED FROM AN ON-CAMPUS POPULATION. ALL STUDENTS WERE EVALUATED FOR PREVIOUS KNOWLEDGE REGARDING ARIMA TECHNIQUES, STATISTICAL BACKGROUND, AND PREVIOUS EXPOSURE TO COMPUTER OPERATION. STUDENTS WERE SELECTED BASED ON: (1.) NO EXPERIENCE WITH ARIMA MODELING, (2.) HAVING TAKEN AT LEAST ONE BASIC STATISTICS COURSE, AND (3.)

HAVING SOME EXPERIENCE WITH COMPUTER TERMINAL OPERATION. THE 30 STUDENTS WERE THEN RANDOMLY ASSIGNED TO ONE OF THREE STATISTICAL SOFTWARE PACKAGES.

THE STATISTICAL SOFTWARE DEVELOPED FOR THIS STUDY WAS WRITTEN IN BASIC LANGUAGE (OPERATING ON A VAX 11/780). THE THREE STATISTICAL SOFTWARE PACKAGES WERE: (1.) SOFTWARE (1.1), WHICH PRODUCED AUCF AND PACF CORRELOGRAMS (A NON-TUTORIAL PACKAGE); (2.) SOFTWARE (1.2), WHICH PRODUCED AUCF AND PACF CORRELOGRAMS PLUS ACCESS TO THEORETICAL ARIMA MODEL CORRELOGRAM EXAMPLES (TUTORIAL A); AND (3.) SOFTWARE (1.3), WHICH PRODUCED AUCF AND PACF CORRELOGRAMS PLUS ACCESS TO A DECISION SUPPORT TUTORIAL (TUTORIAL B). THE FLOW CHARTS OF THESE THREE SOFTWARE PACKAGES ARE PROVIDED IN FIGURES 3A, 3B, AND 3C.

SOFTWARE OPERATION

SOFTWARE PACKAGE (1.1) OPERATED LIKE MOST INTERACTIVE ARIMA PACKAGES, PROVIDING THE USER WITH CORRELOGRAMS AND A DIFFERENCING ROUTINE. THIS WAS CONSIDERED THE "TYPICAL" SOFTWARE APPROACH. PACKAGE (1.2) PROVIDED THE STUDENT WITH THE SAME OPERATION AS IN PACKAGE (1.1), BUT WITH ADDITIONAL ACCESS TO ON-LINE THEORETICAL AUCF AND PACF ARIMA MODEL EXAMPLES. PACKAGE (1.3) PROVIDED THE SAME OPERATION AS IN PACKAGE (1.1), BUT WITH ADDITIONAL ACCESS TO A COMPUTER ASSISTED IDENTIFICATION ROUTINE. THIS IDENTIFICATION ROUTINE EVALUATED BOTH AUCF AND PACF FUNCTIONS FOR ADHERENCE TO THE FOLLOWING NONSEASONAL ARIMA MODEL PATTERNS LISTED IN

TABLE 1.

SINCE THIS STUDY ONLY USED NONSEASONAL UNIVARIATE MODELS, RECOGNITION BY PACKAGE (1.3) WAS BASED ON THE DEGREE OF FIT OF AN ABSOLUTE DECAYING FUNCTION ACCORDING TO SPECIFIED ARIMA MODELS (SEE FIGURE 4). SINCE THE ACTUAL ABSOLUTE DECAY FUNCTIONS WILL NOT EXACTLY FIT THE THEORETICAL DECAY FUNCTIONS , VARIANCE BANDS WERE CREATED AROUND THE THEORETICAL FUNCTIONS. THESE BANDS ALTERNATED BETWEEN BOTH FUNCTIONS AT 5% INCREMENTS UNTIL EITHER A MODEL WAS IDENTIFIED OR 100% WAS REACHED (FIGURE 5). WHILE THE ACTUAL FUNCTIONS SHOULD APPROXIMATE THE THEORETICAL DECAY FUNCTION, THIS WAS SELDOM FOUND EVEN WITH VARIANCE BANDS OF 50%. SUBSEQUENTLY, THE THEORETICAL ABSOLUTE DECAY FUNCTIONS WERE ADJUSTED UPWARD IN 10% INCREMENTS UP TO 80% TO ALLOW FOR DECAY RESEMBLANCE (SEE FIGURE 5).

STUDY DESIGN

A SPLIT-PLOT FACTORIAL DESIGN EMPLOYING ONE BETWEEN-BLOCK TREATMENT (THREE SOFTWARE PACKAGES) AND TWO WITHIN-BLOCK TREATMENTS (FOUR REPEATED TRIALS, AND HIGH AND LOW COMPLEXITY) WAS USED. STUDENTS WERE RANDOMLY ASSIGNED TO THE BETWEEN-BLOCK TREATMENT, PROVIDING 10 STUDENTS PER SOFTWARE PACKAGE. A TOTAL OF 56 TIME-SERIES FILES (100 OBSERVATIONS EACH) WERE THEN PRESENTED FOR IDENTIFICATION, PROVIDING 14 FILES , SEVEN AT EACH COMPLEXITY LEVEL, PER TRIAL PERIOD.

RESULTS

AN ANALYSIS OF VARIANCE WITH REPEATED MEASURES INDICATED A SIGNIFICANT INTERACTING EFFECT BETWEEN COMPLEXITY OF PATTERNS PRESENTED AND TYPE OF SOFTWARE USED (SEE TABLE 2). THE TWO TUTORIAL SOFTWARE PACKAGES, TUTORIAL A (MEAN=3.71) AND TUTORIAL B (MEAN=5.23), SIGNIFICANTLY AIDED RECOGNITION OF ARIMA MODELS OVER THE NON-TUTORIAL PACKAGE (MEAN=2.18). THE FACT THAT EXAMPLES OF THEORETICAL ARIMA MODELS (TUTORIAL A) ENHANCED STUDENT RECOGNITION OVER NO EXAMPLES, IN BOTH HIGH AND LOW COMPLEXITY, DIFFERED FROM A PREVIOUS FINDING (BROWN, 1980). THIS DIFFERENCE WAS ATTRIBUTED TO DIFFERENCES IN STUDENT ABILITY, WITH THE FORMER STUDY USING GRADUATE STUDENTS IN STATISTICS AND THE LATTER STUDY USING NOVICE STUDENTS.

SIMPLE EFFECT TESTS ON THE INTERACTION INDICATED THAT THE RESULTS OF USING THE DECISION SUPPORT TUTORIAL (TUTORIAL B) DID NOT DIFFER SIGNIFICANTLY FROM THE RESULTS OF USING THE EXAMPLE TUTORIAL (TUTORIAL A) WHEN THE COMPLEXITY WAS HIGH, BUT DID PROVIDE A SIGNIFICANT IMPROVEMENT OVER THE EXAMPLES IN THE LOW COMPLEXITY CONDITION (FIGURE 6). IT IS THOUGHT THAT ONE REASON FOR A LACK OF SIGNIFICANT IMPROVEMENT IN THE HIGH COMPLEXITY CONDITION WAS DUE TO THE STUDENTS' INABILITY. THE NOVICE STUDENTS SEEMED TO RELY HEAVILY ON THE DECISION PROVIDED BY TUTORIAL B, EVEN WHEN THE RESULTS WERE CONSIDERED QUESTIONABLE. IT IS SPECULATED THAT THIS FINDING WOULD BE REVERSED GIVEN MORE PROFICIENT USERS.

A SIGNIFICANT NON-INTERACTIVE LEARNING EFFECT WAS ALSO PRESENT OVER THE REPEATED TRIALS.

CONCLUSIONS

THE USE OF ON-LINE TUTORIAL ROUTINES IN THE IDENTIFICATION OF TENTATIVE ARIMA MODELS REDUCES THE NUMBER OF ERRORS MADE BY NOVICE STUDENTS. WHILE THIS SEEMS A REASONABLE STATEMENT, THE TYPE OF TUTORIAL USED IS OF SOME IMPORTANCE. THIS STUDY HAS INDICATED THAT THE USE OF DECISION SUPPORT TUTORIALS THAT PROVIDE THE USER WITH A BEST GUESS ARE NOT APPROPRIATE FOR NOVICE USERS IN COMPLEX SITUATIONS, SINCE THE USERS SEEM TO ACCEPT THE SOFTWARE'S RESULTS AS ALWAYS APPROPRIATE. THE USE OF A SIMPLE EXAMPLE TUTORIAL PROVIDES SIMILAR EFFECTIVENESS IN REDUCING MODEL RECOGNITION ERROR AND DOES NOT PROVIDE THE OPPORTUNITY TO DIRECTLY BIAS THE USER.

STATISTICAL TUTORIALS DO SEEM TO BE AN EFFECTIVE WAY OF ENHANCING THE RECOGNITION OF APPROPRIATE ARIMA MODELS ON A SHORT TERM BASIS. FURTHER STUDY WILL BE NECESSARY TO INVESTIGATE THE EFFECTIVENESS OF VARIOUS STATISTICAL TUTORIALS ON LONG TERM MODEL RECOGNITION.

REFERENCES

BOWERMAN, BL., & O'CONNELL, R.T. EORECASIING AND IIME=SERIES. NOR:H SCITUATE, MASSACHUSETTS: DUXBURY PRESS, 1979.

BOX, G.E.P., AND JENKINS, G.M. IIME=SERIES ANALYSIS EORECASIING AND CONIROL. SAN FRANCISCO: HOLDEN-DAY, 1970.

BROWN, R.L. SOFTWARE PSYCHOLOGY AND THE DATA ANALYTIC SYSTEM: COMPUTER ASSISTED IDENTIFICATION OF CORRELOGRAMS IN TIME-SERIES DATA ANALYSIS. PROCEEDINGS OE IHE EOURIH ANNUAL SPSS USERS AND COORDINAIORS CONEERENCE, 1980, 109-119. (SUMMARY)

MC CLEARY, R., & HAY, R.A. APPLIED IIME=SERIES ANALYSIS EOR IHE SOCIAL SCIENCES. BEVERLY HILLS: SAGE PUBLICATIONS, 1980.

FIGURE 1

ITERATIVE MODEL BUILDING APPROACH

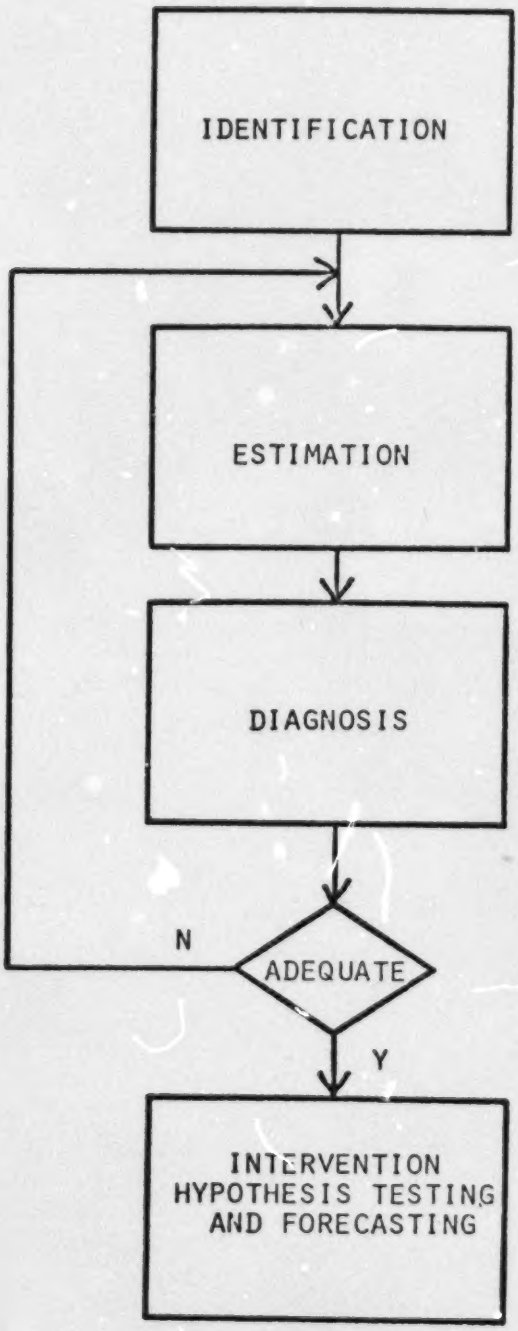


FIGURE 2

AUCF PLOT
MODEL DIFFERENCE = 0

LAG	AUCF	t-Value
1	0.680	4.807
2	0.617	3.147
3	0.540	2.331
4	0.523	2.045
5	0.570	2.062
6	0.464	1.551
7	0.503	1.609
8	0.415	1.263
9	0.401	1.182
10	0.349	1.002
11	0.304	0.856
12	0.246	0.683
13	0.182	0.501
14	0.198	0.542
15	0.094	0.255

PACF PLOT
MODEL DIFFERENCE = 0

LAG	PACF	t-Value
1	0.680	4.807
2	0.289	2.041
3	0.092	0.652
4	0.128	0.907
5	0.237	1.679
6	-.102	-0.719
7	0.144	1.020
8	-.072	-0.506
9	-.005	-0.032
10	-.074	-0.520
11	-.018	-0.127
12	-.156	-1.102
13	-.050	-0.355
14	0.026	0.185
15	-.161	-1.138

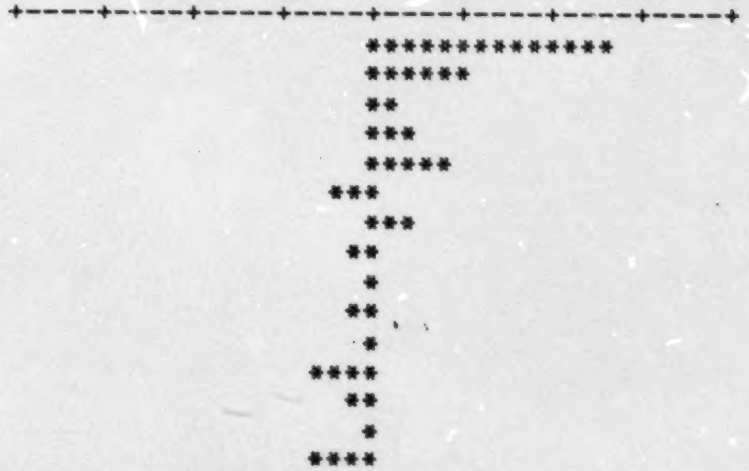


FIGURE 3A

SOFTWARE PACKAGE (1.1)

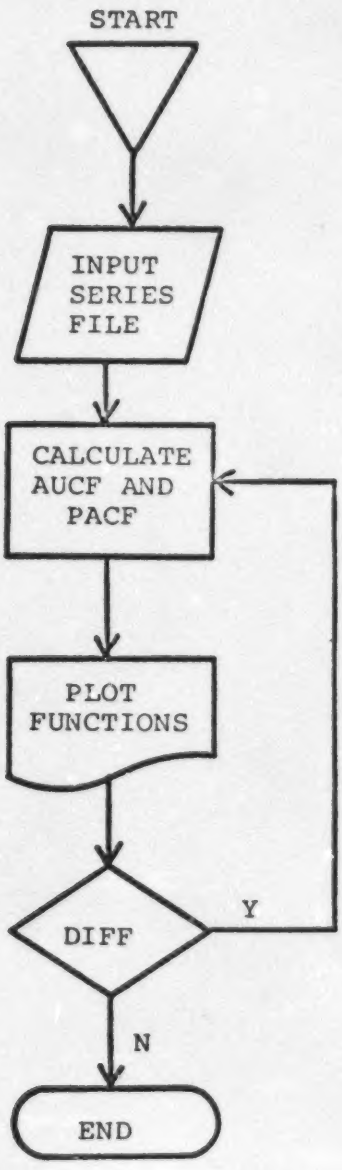


FIGURE 3B

SOFTWARE PACKAGE (1.2)

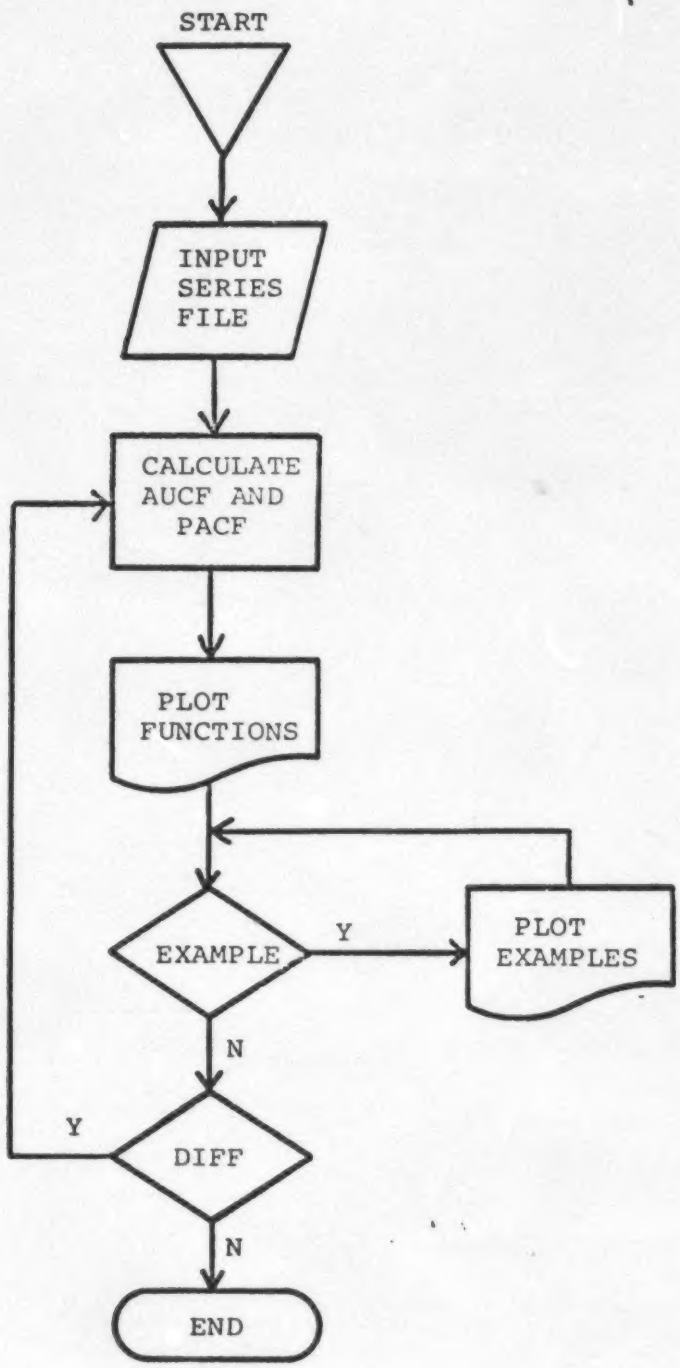


FIGURE 3C

SOFTWARE PACKAGE (1.3)

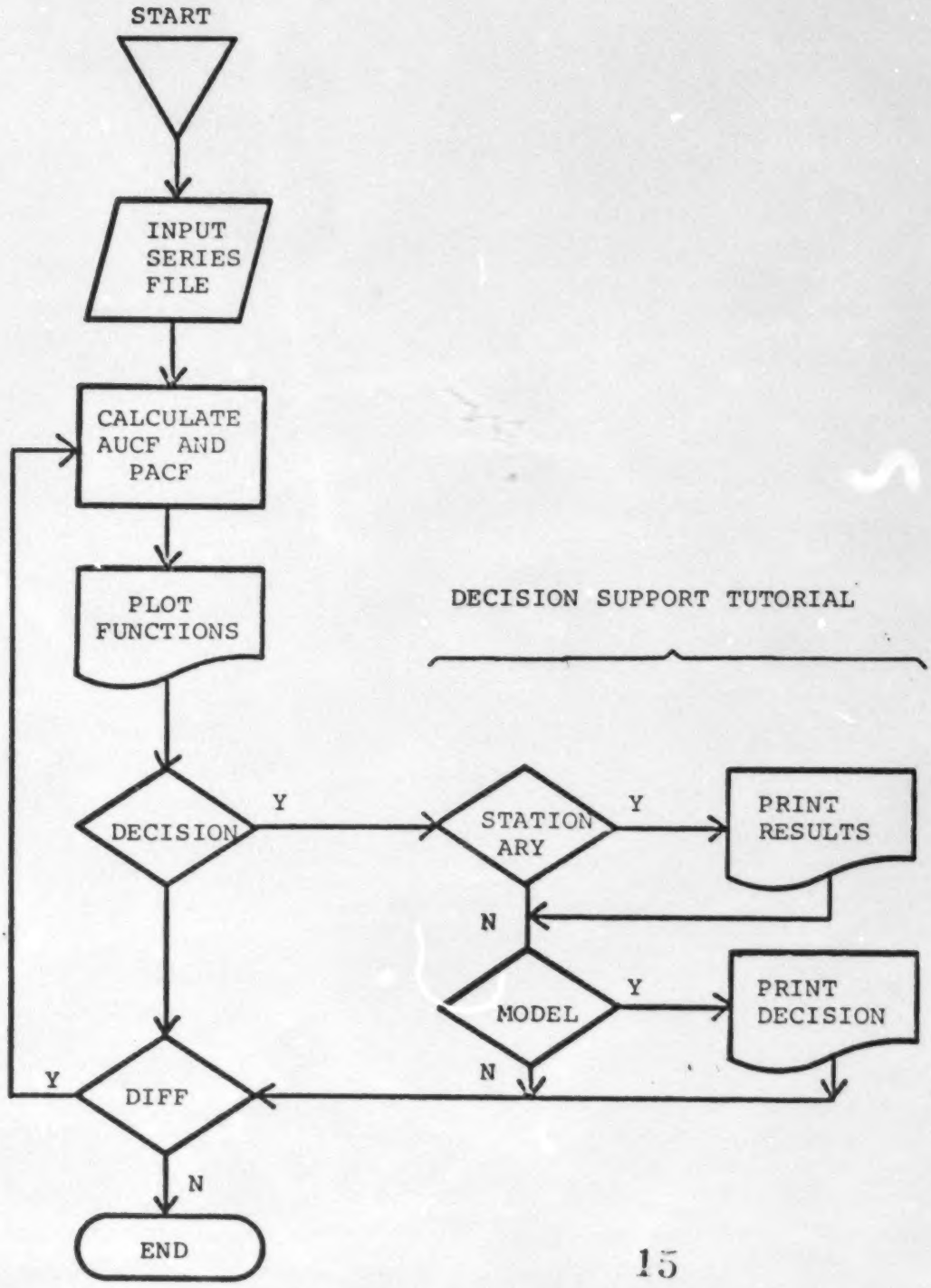


TABLE 1

NONSEASONAL UNIVARIATE ARIMA MODELS					
	WHITE NOISE	AR(1)	AR(2)	MA(1)	MA(2)
AUCF	ALL LAG T-VALUES < 2.00	EXPONENTIAL AND/OR SINE WAVE DECAY	EXPONENTIAL AND/OR SINE WAVE DECAY	LAG 1 T-VALUE > 2.00	LAGS 1 & 2 T-VALUES > 2.00
PACF	ALL LAG T-VALUES < 2.00	LAG 1 T-VALUE > 2.00	LAGS 1 & 2 T-VALUES > 2.00	EXPONENTIAL AND/OR SINE WAVE DECAY	EXPONENTIAL AND/OR SINE WAVE DECAY

FIGURE 4

ABSOLUTE DECAY FUNCTION (FIRST 8 LAGS)

SOFTWARE (1.3) GENERATED DECAY FUNCTION WITH A 20% VARIANCE BAND

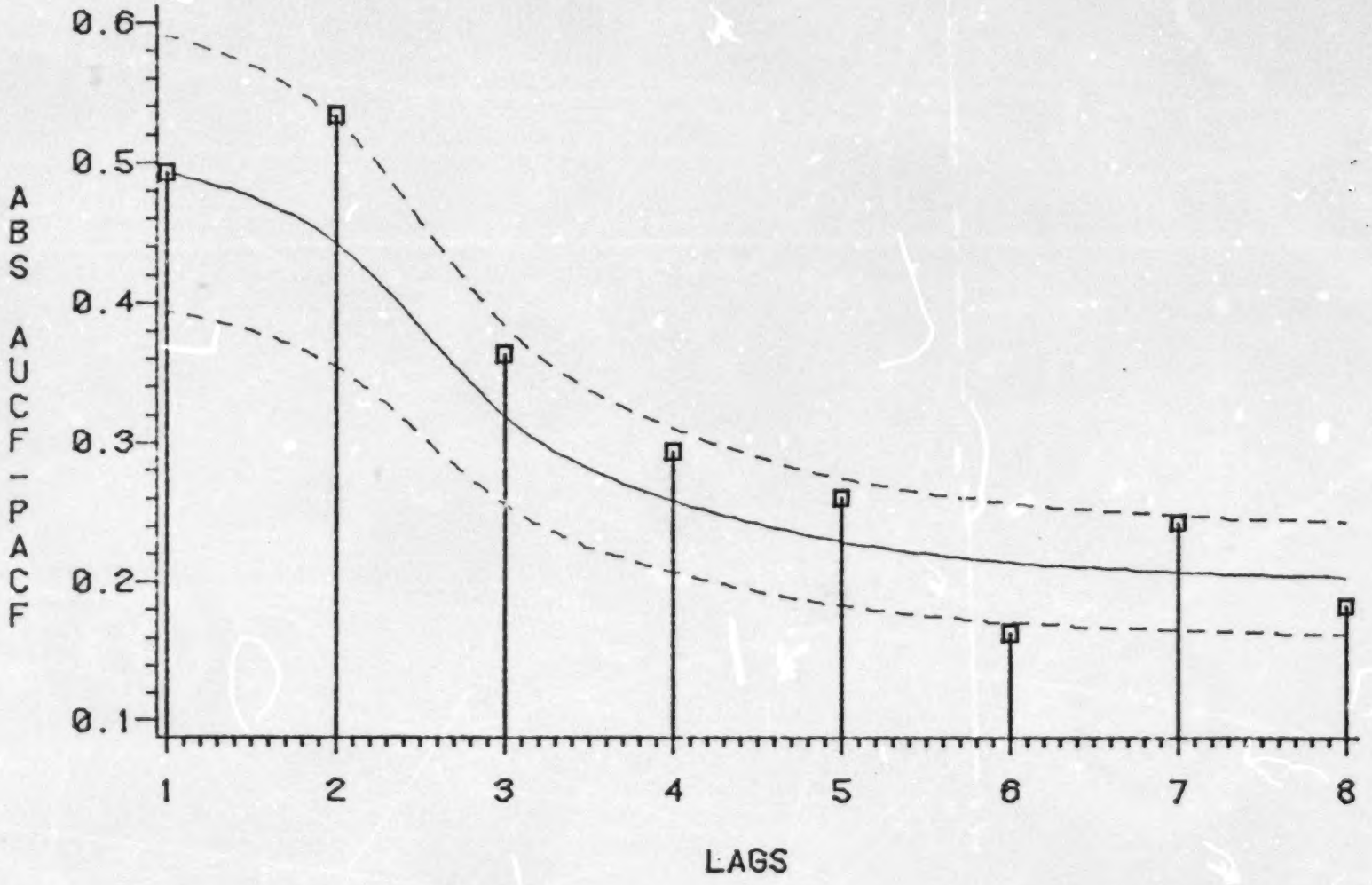
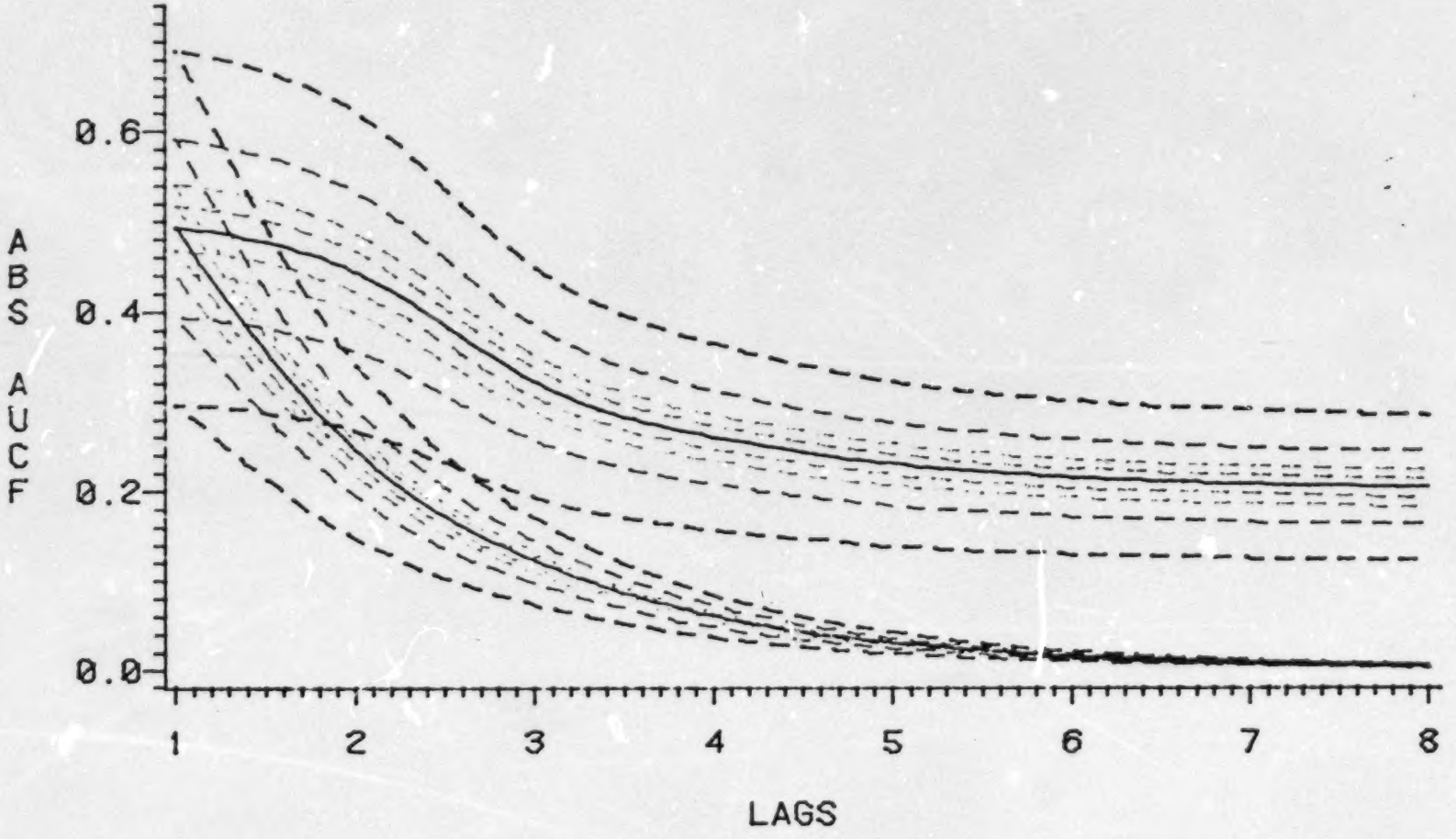


FIGURE 5

ARIMA MODEL IDENTIFICATION USING SOFTWARE (1.3)
ABSOLUTE DECAY FUNCTIONS WITH VARIANCE BANDS



BLUE=5% GREEN=10% RED=20% BLACK=40%

LOWER BLK CURVE=ORIGINAL FUNCTION - UPPER BLK CURVE=80% RISE

TABLE 2

ANALYSIS OF VARIANCE SUMMARY TABLE

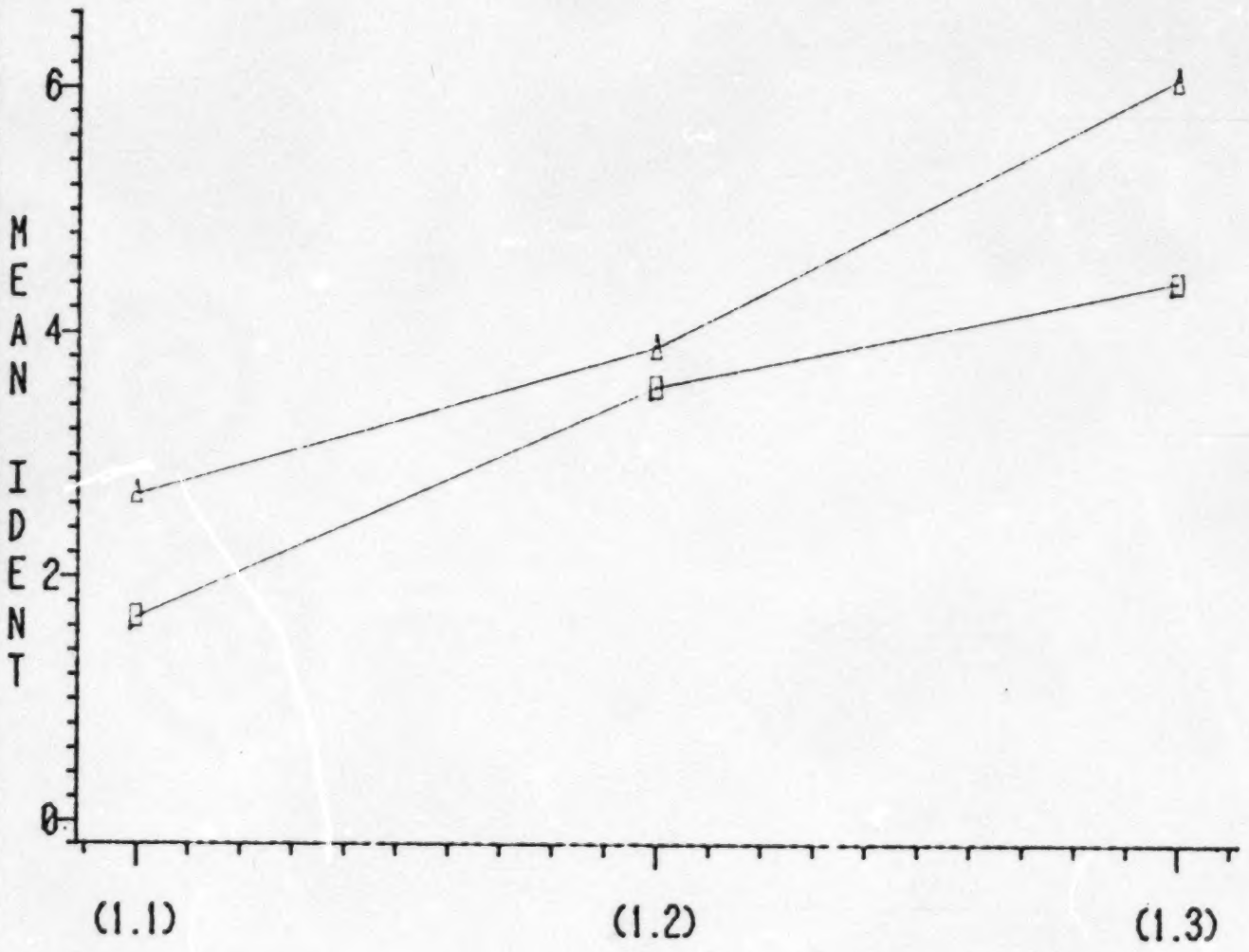
SOURCE	SUM OF SQUARES	MEAN SQUARES	DF	F-RATIO
BETWEEN SUBJECTS				
SOFTWARE (A)	372.11	186.05	2	99.95 *
ERROR	50.26	1.86	27	
WITHIN SUBJECTS				
TRIALS (J)	35.78	11.93	3	14.96 *
AJ	10.76	1.79	6	2.25
ERROR	64.59	0.80	81	
COMPLEXITY (K)	59.00	59.00	1	78.43 *
AK	17.56	8.78	2	11.67 *
ERROR	20.31	0.75	27	
JK	0.98	0.33	3	<1.0
AJK	3.91	0.65	6	1.04
ERROR	50.74	0.63	81	

(*) P < .05

FIGURE 6

TYPE OF SOFTWARE BY LEVEL OF COMPLEXITY

INTERACTION PLOT



BLUE SQUARE=HIGH COMPLEXITY CONDITION

GREEN TRIANGLE=LOW COMPLEXITY CONDITION